ISO/IEC JTC1/SC2/WG2 N4283
L2/12-214
2012-06-20

Title: Preliminary Proposal to Encode the Rohingya Script
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2012-06-20

1 Introduction

This is a preliminary proposal to encode the Rohingya script in the Universal Character Set (ISO/IEC 10646). This document provides a brief description of the writing system, a code chart and names list, character data, and a few specimens. The glyphs used in the code chart are sourced from the "Rohingya Gonya Leyka Noories" typeface developed by Muhammad Noor, with some new glyphs produced by the present author. All information presented here is tentative and may change as a result of additional research. A formal proposal to encode the script will be submitted later.

2 Background

The Rohingya script is used for writing Rohingya (ISO 639-3: rhg), an Indo-Aryan language spoken by one million people in Myanmar (Rakhine State) and by two-hundred thousand people in Bangladesh (Cox's Bazaar District). There are four different scripts used for writing Rohingya: Burmese, Arabic, the Latin-based 'Rohingylish', and the script described here, which was developed by Maulana Hanif in the 1980s.

The script is modeled upon the Arabic script and shows the influence of other scripts; however, it is a constructed script and has no genetic affiliation to other scripts.

There is limited information available on the Rohingya script in English. Most of the materials on the script are written in the Rohingya language. The Rohingya Language Committee and the Rohingya Education Board Myanmar have published primers of the script (see figures 3–7). There are also instructional videos available on YouTube (see figures 8–9). Two typefaces were developed for Rohingya by Muhammad Noor (see table 1).

3 Script Details

3.1 Structure

Rohingya is an alphabetic script that is written from right to left. Letters join to following letters at the right edge. The shapes of letters do not change. Vowels are always written.

3.2 Vowels

There is 1 vowel-carrier letter: 𐴀 vowel-carrier letter.

and 5 vowel letters:

AA	I	U	E	O

When written independently, the vowel-carrier letter represents the sound /a/. The vowel letters are never written independently or word-initially. When they occur in such contexts, they are written with the vowel-carrier letter and after the carrier.

3.3 Consonants

There are 28 consonants:

BA	XA	SA	WA
PA	FA	SHA	WWA
TA	DA	KA	YA
TTA	TAH	GA	YYA
JA	RA	LA	NGA
CA	RRA	MA	NYA
HA	ZAH	NA	QA

3.4 Nasalization

The NASAL LETTER is used for marking vowel nasalization. It is called *na konna* in Rohingya and is similar in function to U+06BA ARABIC LETTER NOON GHUNNA.

3.5 Gemination

The ◌ SIGN SHADDA indicates consonant gemination. It is called *tossi* in Rohingya and is similar in function to U+0651 ARABIC SHADDA.

3.6 'Sukun'

The ◌ SIGN SUKUN is written with certain consonant letters to indicate explicitly the absence of a vowel at the end of a word. The sign is also used for writing a letter when it occurs independently. It is called *sakin* in Rohingya and is similar in function to U+0652 ARABIC SUKUN.

In Rohingya, certain bare consonants are not represented using the SUKUN. See, for example, the isolated forms of the letters in figure 3.

Bare forms of certain letters are written using distinct glyphs: /m/ is represented as instead of as <MA, ◌ SIGN SUKUN>; /l/ is represented as in some texts instead of as <LA, ◌ SIGN SUKUN>.

3.7 Tonal Signs

1. ◌𐴤 TONE-1 This sign indicates short high tone. It is called *haar bay* in Rohingya. This tonal sign corresponds to U+08EA ARABIC TONE ONE DOT ABOVE and U+08ED ARABIC TONE ONE DOT BELOW.
2. ◌𐴥 TONE-2 This sign indicates long falling tone. It is called *thela* in Rohingya. This tonal sign corresponds to U+08EB ARABIC TONE TWO DOTS ABOVE and U+08EE ARABIC TONE TWO DOTS BELOW.
3. ◌𐴦 TONE-3 This sign indicates long rising tone. It is called *thana* in Rohingya. This tonal sign corresponds to U+08EC ARABIC TONE LOOP ABOVE and U+08EF ARABIC TONE LOOP BELOW.

3.8 Digits

There is a full set of decimal digits: 𐴰 ZERO, 𐴱 ONE, 𐴲 TWO, 𐴳 THREE, 𐴴 FOUR, 𐴵 FIVE, 𐴶 SIX, 𐴷 SEVEN, 𐴸 EIGHT, 𐴹 NINE. Similar to Arabic, the Rohingya digits are written from left to right. The Arabic style ٠ is attested as a glyphic variant for 𐴰 ZERO.

3.9 Punctuation

There is no script-specific punctuation. The U+002E FULL STOP is commonly used, as are Arabic signs, such as the ، U+060C ARABIC COMMA and ؟ U+061F ARABIC QUESTION MARK.

The ـ TATWEEL is used for graphical elongation or justification. It is similar to the corresponding character U+0640 ARABIC TATWEEL.

4 Preliminary Character Properties

4.1 Unicode Character Data

```
1EA00;ROHINGYA VOWEL-CARRIER LETTER;Lo;0;R;;;;;N;;;;;
1EA01;ROHINGYA LETTER BA;Lo;0;R;;;;;N;;;;;
1EA02;ROHINGYA LETTER PA;Lo;0;R;;;;;N;;;;;
1EA03;ROHINGYA LETTER TA;Lo;0;R;;;;;N;;;;;
1EA04;ROHINGYA LETTER TTA;Lo;0;R;;;;;N;;;;;
1EA05;ROHINGYA LETTER JA;Lo;0;R;;;;;N;;;;;
1EA06;ROHINGYA LETTER CA;Lo;0;R;;;;;N;;;;;
1EA07;ROHINGYA LETTER HA;Lo;0;R;;;;;N;;;;;
1EA08;ROHINGYA LETTER XA;Lo;0;R;;;;;N;;;;;
1EA09;ROHINGYA LETTER FA;Lo;0;R;;;;;N;;;;;
1EA0A;ROHINGYA LETTER DA;Lo;0;R;;;;;N;;;;;
1EA0B;ROHINGYA LETTER TAH;Lo;0;R;;;;;N;;;;;
1EA0C;ROHINGYA LETTER RA;Lo;0;R;;;;;N;;;;;
1EA0D;ROHINGYA LETTER RRA;Lo;0;R;;;;;N;;;;;
1EA0E;ROHINGYA LETTER ZAH;Lo;0;R;;;;;N;;;;;
1EA0F;ROHINGYA LETTER SA;Lo;0;R;;;;;N;;;;;
1EA10;ROHINGYA LETTER SHA;Lo;0;R;;;;;N;;;;;
1EA11;ROHINGYA LETTER KA;Lo;0;R;;;;;N;;;;;
1EA12;ROHINGYA LETTER GA;Lo;0;R;;;;;N;;;;;
1EA13;ROHINGYA LETTER LA;Lo;0;R;;;;;N;;;;;
1EA14;ROHINGYA LETTER MA;Lo;0;R;;;;;N;;;;;
1EA15;ROHINGYA LETTER NA;Lo;0;R;;;;;N;;;;;
1EA16;ROHINGYA LETTER WA;Lo;0;R;;;;;N;;;;;
```

```
1EA17;ROHINGYA LETTER WWA;Lo;0;R;;;;;N;;;;;
1EA18;ROHINGYA LETTER YA;Lo;0;R;;;;;N;;;;;
1EA19;ROHINGYA LETTER YYA;Lo;0;R;;;;;N;;;;;
1EA1A;ROHINGYA LETTER NGA;Lo;0;R;;;;;N;;;;;
1EA1B;ROHINGYA LETTER NYA;Lo;0;R;;;;;N;;;;;
1EA1C;ROHINGYA LETTER QA;Lo;0;R;;;;;N;;;;;
1EA1D;ROHINGYA VOWEL LETTER AA;Lo;0;R;;;;;N;;;;;
1EA1E;ROHINGYA VOWEL LETTER I;Lo;0;R;;;;;N;;;;;
1EA1F;ROHINGYA VOWEL LETTER U;Lo;0;R;;;;;N;;;;;
1EA20;ROHINGYA VOWEL LETTER E;Lo;0;R;;;;;N;;;;;
1EA21;ROHINGYA VOWEL LETTER O;Lo;0;R;;;;;N;;;;;
1EA22;ROHINGYA NASAL LETTER;Lo;0;R;;;;;N;;;;;
1EA23;ROHINGYA SIGN TONE-1;Mn;220;NSM;;;;;N;;;;;
1EA24;ROHINGYA SIGN TONE-2;Mn;220;NSM;;;;;N;;;;;
1EA25;ROHINGYA SIGN TONE-3;Mn;220;NSM;;;;;N;;;;;
1EA26;ROHINGYA SIGN SHADDA;Mn;33;NSM;;;;;N;;;;;
1EA27;ROHINGYA SIGN SUKUN;Mn;34;NSM;;;;;N;;;;;
1EA28;ROHINGYA TATWEEL;Lm;0;AL;;;;;N;;;;;
1EA30;ROHINGYA DIGIT ZERO;Nd;0;AN;;0;0;0;N;;;;;
1EA31;ROHINGYA DIGIT ONE;Nd;0;AN;;1;1;1;N;;;;;
1EA32;ROHINGYA DIGIT TWO;Nd;0;AN;;2;2;2;N;;;;;
1EA33;ROHINGYA DIGIT THREE;Nd;0;AN;;3;3;3;N;;;;;
1EA34;ROHINGYA DIGIT FOUR;Nd;0;AN;;4;4;4;N;;;;;
1EA35;ROHINGYA DIGIT FIVE;Nd;0;AN;;5;5;5;N;;;;;
1EA36;ROHINGYA DIGIT SIX;Nd;0;AN;;6;6;6;N;;;;;
1EA37;ROHINGYA DIGIT SEVEN;Nd;0;AN;;7;7;7;N;;;;;
1EA38;ROHINGYA DIGIT EIGHT;Nd;0;AN;;8;8;8;N;;;;;
1EA39;ROHINGYA DIGIT NINE;Nd;0;AN;;9;9;9;N;;;;;
```

4.2 Arabic Shaping Data

```
# Rohingya Characters

1EA00; ROHINGYA VOWEL-CARRIER LETTER; L; No_Joining_Group
1EA01; ROHINGYA LETTER BA; D; No_Joining_Group
...
1EA22; ROHINGYA NASAL LETTER; D; No_Joining_Group
1EA28; ROHINGYA TATWEEL; C; No_Joining_Group
```

5 References

noorismail52. 2011a. “Rohingya mother Language (letters) Part 1.flv”. `http://www.youtube.com/watch?v=w4h6w6NyvOU`

———. 2011b. “Rohingya mother Language (letters) Part 3.flv”. `http://www.youtube.com/watch?v=pylvjjTQG8c`

Rohingya Education Board Myanmar. *Ruhainggya Zubanor Fonna: Hisab* [“Script of the Rohingya language: Counting”]. Ek kelasottu dui kelas: lego ar foro [“From Class 1 to Class 2: Read and Write”].

Rohingya Language Committee. [A]. *Kayda Ruwainggya Zubanor* [Primer of the Rohingya Language].

———. [B]. *Ruwaingya Zubanor Foyla Kitab* [First Book of the Rohingya Language].

6 Acknowledgments

I am thankful to Mattias Persson and Ian James for bringing the Rohingya script to my attention. Lorna Priest introduced me to James Lloyd-Williams, who very generously provided copies of Rohingya primers. The present work would not be possible without these materials.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.

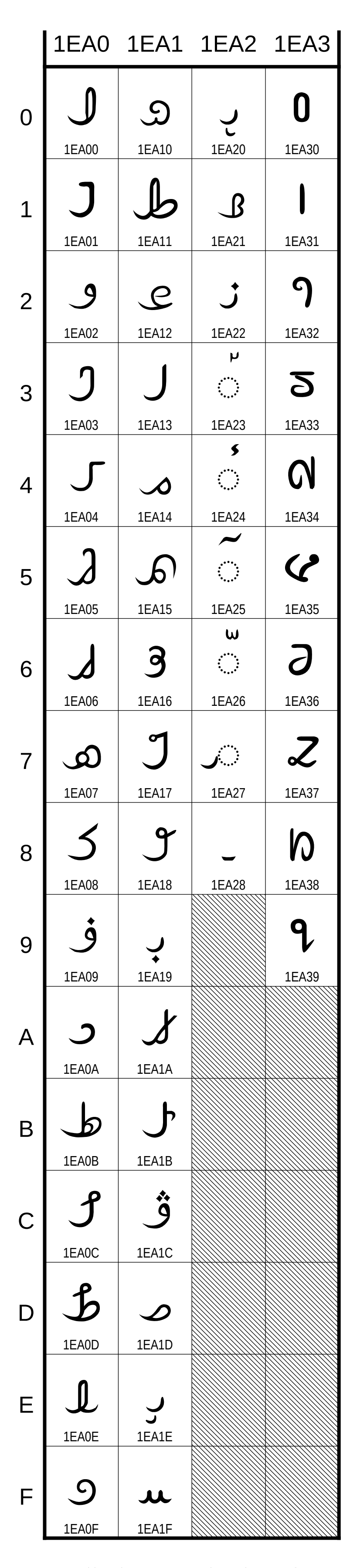


Figure 1: Preliminary code chart for Rohingya.

Vowel-carrier

1EA00 ROHINGYA VOWEL-CARRIER LETTER

Consonants

1EA01 ROHINGYA LETTER BA
1EA02 ROHINGYA LETTER PA
1EA03 ROHINGYA LETTER TA
1EA04 ROHINGYA LETTER TTA
1EA05 ROHINGYA LETTER JA
1EA06 ROHINGYA LETTER CA
1EA07 ROHINGYA LETTER HA
1EA08 ROHINGYA LETTER XA
1EA09 ROHINGYA LETTER FA
1EA0A ROHINGYA LETTER DA
1EA0B ROHINGYA LETTER TAH
1EA0C ROHINGYA LETTER RA
1EA0D ROHINGYA LETTER RRA
1EA0E ROHINGYA LETTER ZAH
1EA0F ROHINGYA LETTER SA
1EA10 ROHINGYA LETTER SHA
1EA11 ROHINGYA LETTER KA
1EA12 ROHINGYA LETTER GA
1EA13 ROHINGYA LETTER LA
1EA14 ROHINGYA LETTER MA
1EA15 ROHINGYA LETTER NA
1EA16 ROHINGYA LETTER WA
1EA17 ROHINGYA LETTER WWA
• kinna wa
1EA18 ROHINGYA LETTER YA
1EA19 ROHINGYA LETTER YYA
• shunyo ya
1EA1A ROHINGYA LETTER NGA
1EA1B ROHINGYA LETTER NYA
1EA1C ROHINGYA LETTER QA

Vowels

1EA1D ROHINGYA VOWEL LETTER AA
1EA1E ROHINGYA VOWEL LETTER I
1EA1F ROHINGYA VOWEL LETTER U
1EA20 ROHINGYA VOWEL LETTER E
1EA21 ROHINGYA VOWEL LETTER O

Nasal letter

1EA22 ROHINGYA NASAL LETTER
= na konna

Tone marks

1EA23 ◌ ROHINGYA SIGN TONE-1
= har bay
• short high tone
1EA24 ◌ ROHINGYA SIGN TONE-2
= thela
• long falling tone
1EA25 ◌ ROHINGYA SIGN TONE-3
= thana
• long rising

Various signs

1EA26 ◌ ROHINGYA SIGN SHADDA
= tossi
• gemination sign
1EA27 ROHINGYA SIGN SUKUN
= sakin
• indicates absence of vowel

Punctuation

1EA28 ROHINGYA TATWEEL

Digits

1EA30 ROHINGYA DIGIT ZERO
1EA31 ROHINGYA DIGIT ONE
1EA32 ROHINGYA DIGIT TWO
1EA33 ROHINGYA DIGIT THREE
1EA34 ROHINGYA DIGIT FOUR
1EA35 ROHINGYA DIGIT FIVE
1EA36 ROHINGYA DIGIT SIX
1EA37 ROHINGYA DIGIT SEVEN
1EA38 ROHINGYA DIGIT EIGHT
1EA39 ROHINGYA DIGIT NINE

Figure 2: Preliminary names list for Rohingya.



Figure 3: Chart of Rohingya script from a hand-written primer (from Ruwainggya Zuban Komiti (A): 1).

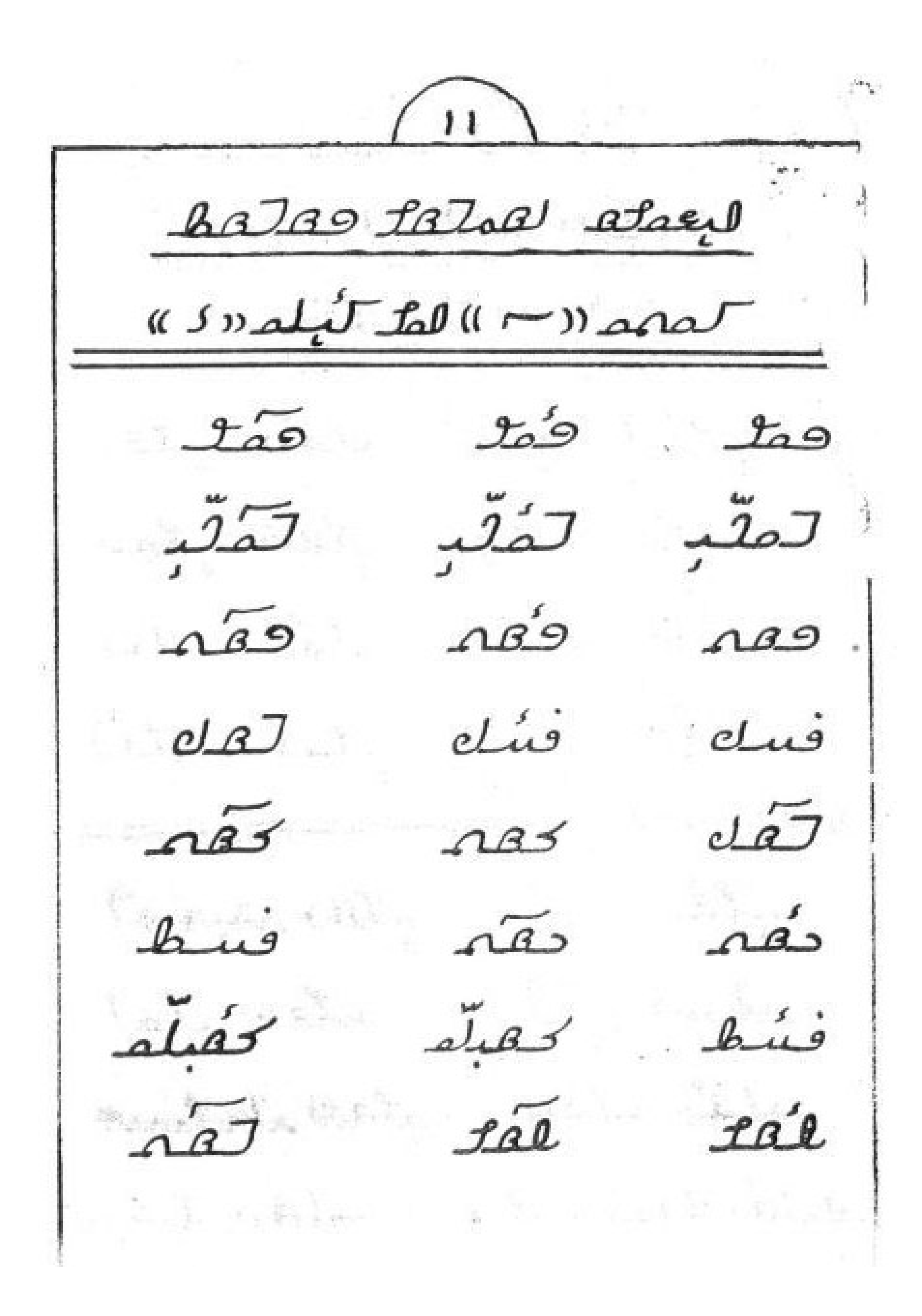


Figure 4: Table showing use of tonal signs from a hand-written primer (from Ruwainggya Zuban Komiti (A): 11).

Figure 5: Example of running Rohingya text from a hand-written primer (from Ruwainggya Zuban Komiti (B): 1).

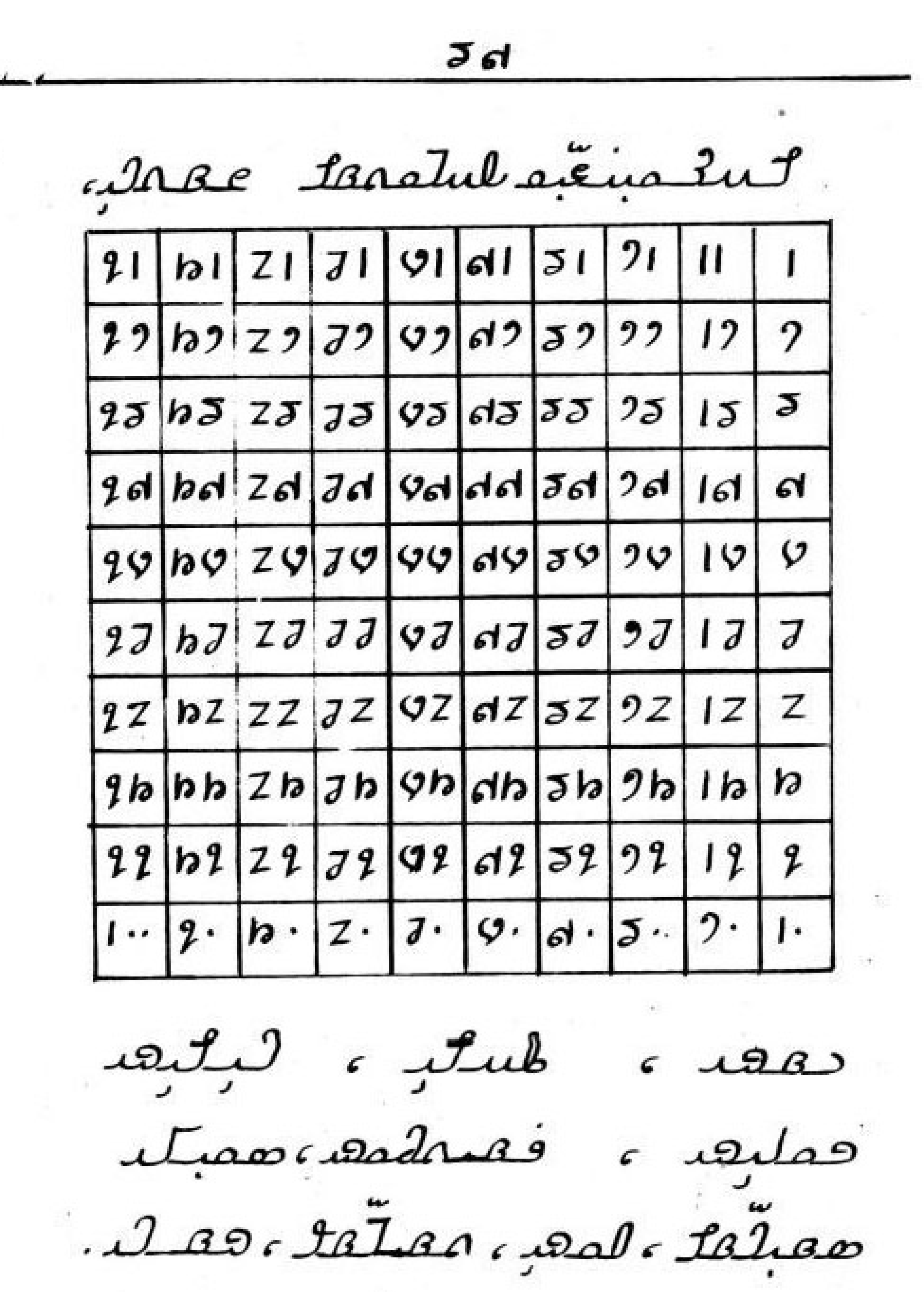


Figure 6: Chart of digits from a hand-written primer (from Ruwainggya Zuban Komiti (B): 34).

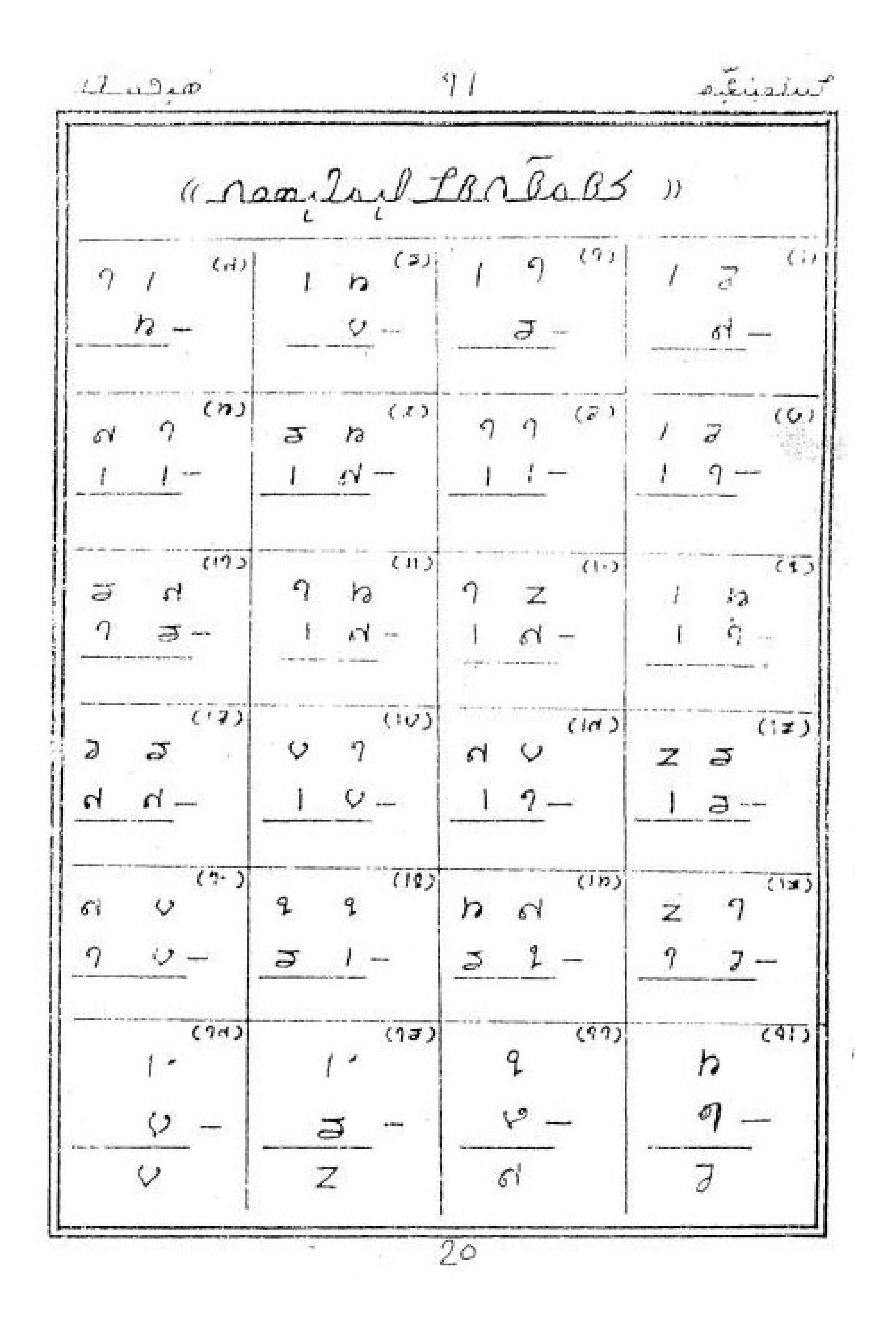


Figure 7: Excerpt from an primary-level arithmetic book written in Rohingya (from Ruwainggya Education Board Myanmar: 21).



Figure 8: Use of a Rohingya typeface in a digital video (from noorismail52 2011a: frame 3).



Figure 9: Use of a Rohingya typeface in a digital video (from noorismail52 2011b: frame 103).

	G	K		G	K		G	K
VOW.CR	[illegible]	[illegible]	KA	[illegible]	[illegible]	NASAL	[illegible]	[illegible]
BA	[illegible]	[illegible]	GA	[illegible]	[illegible]	TONE-1	[illegible]	[illegible]
PA	[illegible]	[illegible]	LA	[illegible]	[illegible]	TONE-2	[illegible]	[illegible]
TA	[illegible]	[illegible]	MA	[illegible]	[illegible]	TONE-3	[illegible]	[illegible]
TTA	[illegible]	[illegible]	NA	[illegible]	[illegible]	SHADDA	[illegible]	[illegible]
JA	[illegible]	[illegible]	WA	[illegible]	[illegible]	SUKUN	[illegible]	[illegible]
CA	[illegible]	[illegible]	WWA	[illegible]	[illegible]	TATWEEL	[illegible]	[illegible]
HA	[illegible]	[illegible]	YA	[illegible]	[illegible]	ZERO	[illegible]	[illegible]
XA	[illegible]	[illegible]	YYA	[illegible]	[illegible]	ONE	[illegible]	[illegible]
FA	[illegible]	[illegible]	NGA	[illegible]	[illegible]	TWO	[illegible]	[illegible]
DA	[illegible]	[illegible]	NYA	[illegible]	[illegible]	THREE	[illegible]	[illegible]
TAH	[illegible]	[illegible]	QA	[illegible]	[illegible]	FOUR	[illegible]	[illegible]
RA	[illegible]	[illegible]	AA	[illegible]	[illegible]	FIVE	[illegible]	[illegible]
RRA	[illegible]	[illegible]	I	[illegible]	[illegible]	SIX	[illegible]	[illegible]
ZAH	[illegible]	[illegible]	U	[illegible]	[illegible]	SEVEN	[illegible]	[illegible]
SA	[illegible]	[illegible]	E	[illegible]	[illegible]	EIGHT	[illegible]	[illegible]
SHA	[illegible]	[illegible]	O	[illegible]	[illegible]	NINE	[illegible]	[illegible]

Table 1: Comparison of digitized Rohingya typefaces: "Rohingya Gonya Leyka Noories" ('G') and "Rohingya Kuna Leyka Noories" ('K') designed by Muhammad Noor.